

基于图卷积神经网络的交通事故风险预测系统

一. **动机：**现今主流基于深度学习的交通事故风险预测系统都是将区域网格化，然后采用深度学习框架进行预测。如图 1 所示。而实际上，交通事故多发生在道路上，城市的路网结构实际是一个网状结构而不是平面结构。因此对整个区域进行网格化之后在进行预测是不合理的，不能直接的告诉交警和司机哪条道路需要绕行或需要治理。

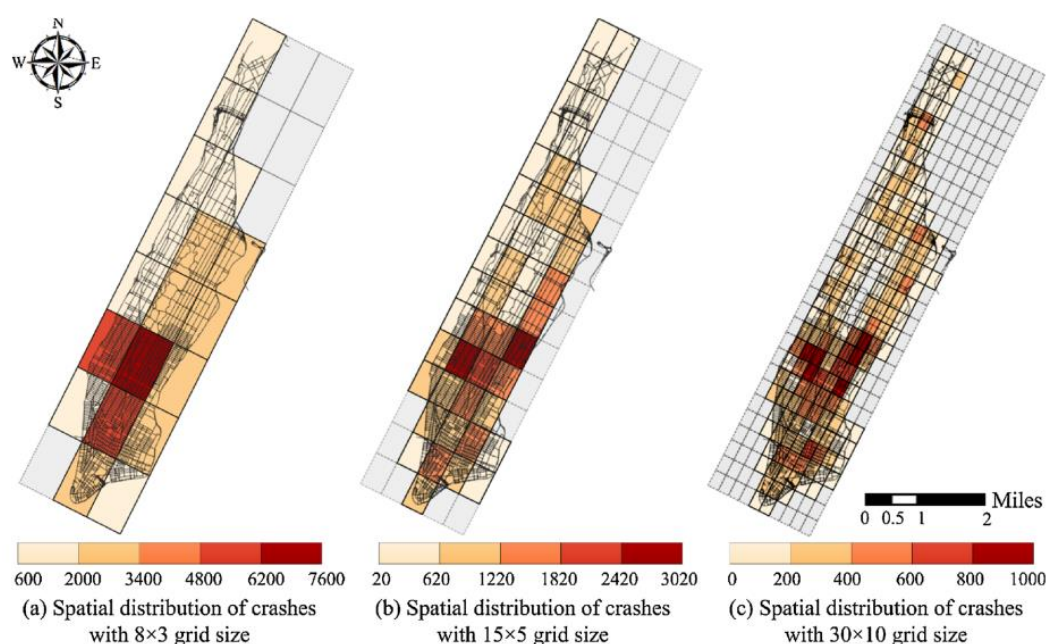


Fig. 1. Spatial distribution of crashes in Manhattan area with different grid sizes.

路网是由路段和路段间的交叉点（路口）构成，因此，路网可以近似的看成是一个图结构，可以从图的角度来对路网进行。

图卷积神经网络（graph convolutional neural network，GCNN）作为近几年提出的一种新技术，可以认为是卷积神经网络（CNN）的一种泛化，不仅可以处理 Euclidean domains 的数据（图像（二维）、视频（一维）），还可以处理非 Euclidean domains 的数据（图结构，三维流体结构）。将路网结构近似为一张图结构，那么可以采用图卷积神经网络来预测交通事故发生风险。

二. **相关工作：**据我们所知，目前图卷积神经网络在交通事故预测方向的工作还没有，在交通领域的研究也比较少。Lin L 等人利用图卷积神经网络和 LSTM 模型相结合来预测美国曼哈顿地区共享单车站点每小时的需求量。

Lin L , He Z , Peeta S . Predicting Station-level Hourly Demands in a Large-scale Bike-sharing Network: A Graph Convolutional Neural Network Approach[J]. Transportation Research Part C , 2018.

美国曼哈顿区域的共享单车具有固定的取车与停车点，论文采用不同的指标，在共享单车站点之间建立图结构，图 2 展示共享单车的站点展示及相应的图结构。

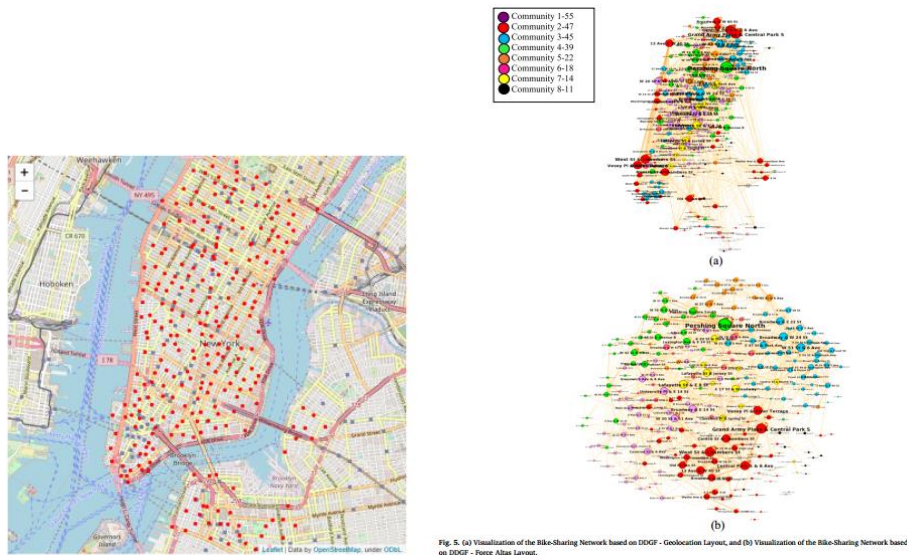


Fig. 5. (a) Visualization of the Bike-Sharing Network based on DDGF - Geolocation Layout, and (b) Visualization of the Bike-Sharing Network based on DDGF - Force Atlas Layout.

图 2 共享单车站点与相应的图结构

论文利用图卷积神经网络，寻找站点之间的特征，从而构建了两种基于卷积神经网络的预测模型，两种模型分别如图三所示。

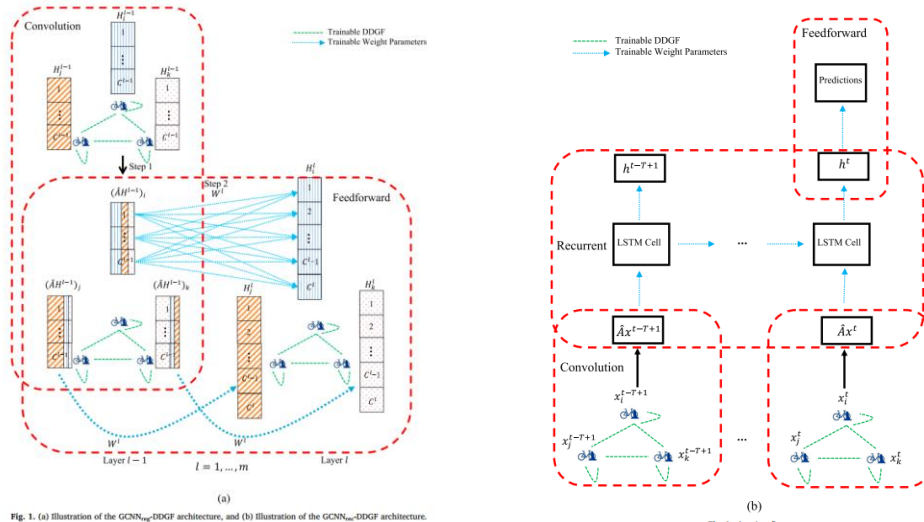


Fig. 1. (a) Illustration of the GCNN-DDGF architecture, and (b) Illustration of the GCNN-DDGF architecture.

Fig. 1. (continued)

图 3. 共享单车站点需求量的卷积神经网络预测模型

三. 工作目标：根据历史交通事故数据，预测某条路段未来一小时、一天、甚至一周的交通事故发生情况。

预测的路段就是自然的路段，及路网结构中由路口拆分出的路段。预测的时间范围，可根据训练模型时采用的最小时间粒度确定。

四. 工作方案：基于图卷积神经网络的交通事故风险预测系统的工作的步骤分三步进行。

1. 路网转化为图结构

在图卷积神经网络中，大部分的操作都是对顶点进行预测，因此，需要转为为以边为顶点的预测。因此可以将路网结构中路段作为顶点，路口最为边，转化为路网的对偶图。图 4 展示了一个简单的示例。

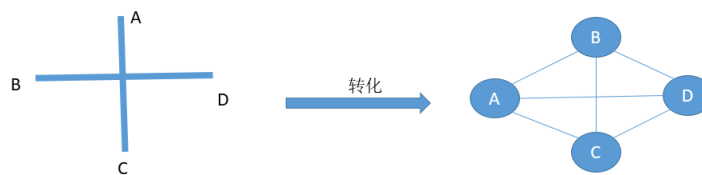


图 4 路网转为为图简单示例

如图 4 所示，AC、BD 为两条道路交叉道路，将道路分为四段分别为 A、B、C、D，则可将 A、B、C、D 作为四个点，将他们的交叉路口作为边，形成图中的图网络。在实际计算过程中，若路段长度过长，可近似在中间加入新的路口，将一段分为两段。

路网结构是很复杂的结构，有许多条道路，在路网转化为图的时候不需要考虑所有的道路。在实际实验过程中，可以选取一块很小的区域（例如：合肥市可以选取一环以内的区域），统计区域内所有道路的事故数，将道路事故发生数目较少的道路排除，仅保留道路事故发生数多的道路，使用这些道路构建路网，并对偶的转化为图结构。

2. 预测模型的构建与训练

采用 GCNN 模型寻找路网结构的特征，并将特征作为 LSTM 或者随机决策树等模型中进行预测。具体实现方式可参考 Lin L 等人的论文。

图中顶点的值，可确定为该路段单位时间内发生的交通事故数目，若有其他数据也可以个构建其他交通事故记录数据。

五. 模型评价：

针对图样的数据，可以直接采用 SVM 等其他分类预测模型进行测试。可以与网格化的深度学习方法进行比较。（这里的可比较性存疑，因为区域网格化和道路图结构是构建的是两种不同类型的数据集，网格化是针对的区域而无法确切的预测道路的事故情况）

模型评价公式，采用基本评价公式：

$$MAE = \frac{1}{n} \sum_{i=1}^n |r_i - \hat{r}_i|$$

$$MRE = \frac{1}{n} \sum_{i=1}^n \frac{|r_i - \hat{r}_i|}{r_i}$$

$$RMSE = \frac{1}{n} \left[\sum_{i=1}^n (r_i - \hat{r}_i)^2 \right]^{\frac{1}{2}}$$

六. 拓展：根据陈为老师建议，该论文投可视化方面期刊，会加入可视化展示界面，若能做成系统，则可以形成 user study，作为论文的一部分。

七. 创新点：

1. 将图卷积神经网络引入到交通事故预测中来
 2. 可以预测某条路段的交通事故发生情况，是规避事故多发路段和治理交通事故有了更明确的区域
 3. 构建的基于图卷积神经网络的交通事故风险预测的可视化界面及系统
- 特色：以图卷积神经网络作为模型内核，以可视化展示界面为特色，以事故预测精度为目标

八. 论文阅读

Kipf T N , Welling M . Semi-Supervised Classification with Graph Convolutional Networks[J]. ICLR 2017.

Zhou J , Cui G , Zhang Z , et al. Graph Neural Networks: A Review of Methods and Applications[J]. 2018.

Defferrard, Michaël, Bresson X , Vandergheynst P . Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering[J]. 2016.